ORIGINAL RESEARCH　　　　　　　　　　　　　　　　　**Open Access**

In silico drug metabolism and pharmacokinetic profiles of natural products from medicinal plants in the Congo basin

Fidele Ntie-Kang[1,2,3*], Lydia L Lifongo[2], James A Mbah[4], Luc C Owono Owono[1,5], Eugene Megnassan[6], Luc Meva'a Mbaze[7], Philip N Judson[8], Wolfgang Sippl[3] and Simon M N Efange[4]

Abstract

Purpose: Drug metabolism and pharmacokinetics (DMPK) assessment has come to occupy a place of interest during the early stages of drug discovery today. The use of computer modelling to predict the DMPK and toxicity properties of a natural product library derived from medicinal plants from Central Africa (named ConMedNP). Material from some of the plant sources are currently employed in African Traditional Medicine.

Methods: Computer-based methods are slowly gaining ground in this area and are often used as preliminary criteria for the elimination of compounds likely to present uninteresting pharmacokinetic profiles and unacceptable levels of toxicity from the list of potential drug candidates, hence cutting down the cost of discovery of a drug. In the present study, we present an *in silico* assessment of the DMPK and toxicity profile of a natural product library containing ~3,200 compounds, derived from 379 species of medicinal plants from 10 countries in the Congo Basin forests and savannas, which have been published in the literature. In this analysis, we have used 46 computed physico-chemical properties or molecular descriptors to predict the absorption, distribution, metabolism and elimination and toxicity (ADMET) of the compounds.

Results: This survey demonstrated that about 45% of the compounds within the ConMedNP compound library are compliant, having properties which fall within the range of ADME properties of 95% of currently known drugs, while about 69% of the compounds have ≤ 2 violations. Moreover, about 73% of the compounds within the corresponding "drug-like" subset showed compliance.

Conclusions: In addition to the verified levels of "drug-likeness", diversity and the wide range of measured biological activities, the compounds from medicinal plants in Central Africa show interesting DMPK profiles and hence could represent an important starting point for hit/lead discovery.

Keywords: ADMET; Drug discovery; Descriptors; *In silico*; Medicinal plants; Natural products

Background

Natural products (NPs) have always played an important role in drug discovery until today (Li and Vederas 2009; Chin et al. 2006; Newman 2008; Harvey 2008; Koehn and Carter 2005). This is because they both serve as active principles in drugs and as templates for the synthesis of new drugs (Newman 2008; Efange 2002). Additionally, a

good proportion of drugs which have been approved for clinical trials, are either NPs or their analogues (Butler 2005). What makes NPs unique is that they are often rich in stereogenic centres and cover segments of chemical space which are typically not occupied by a majority of their synthetic counterparts (Wetzel et al. 2007; Grabowski et al. 2008). Moreover, NPs generally contain more oxygen atoms and less aromatic atoms on average, when compared with "drug-like" molecules (Grabowski and Schneider 2007). They sometimes fail the test for "drug-likeness" due to the fact

---

* Correspondence: ntiekfidele@gmail.com
[1]CEPAMOQ, Faculty of Science, University of Douala, P.O. Box 8580, Douala, Cameroon
[2]Chemical and Bioactivity Information Centre, Department of Chemistry, Faculty of Science, University of Buea, P.O. Box 63, Buea, Cameroon
Full list of author information is available at the end of the article

that they are often bulkier than synthetic drugs (Quinn et al. 2008).

The fact that more and more drugs fail to enter the market as a result of poor pharmacokinetic profiles, has necessitated the inclusion of pharmacokinetic considerations at earlier stages of drug discovery programs (Hodgson 2001; Navia and Chaturvedi 1996). This requires the search for lead compounds which can be easily orally absorbed, easily transported to their desired site of action, not easily metabolised into toxic products before reaching the targeted site of action and easily eliminated from the body before accumulating in sufficient amounts that may produce adverse side effects. The sum of the above mentioned properties is often referred to as ADME (absorption, distribution, metabolism and elimination) properties, or better still ADMET, ADME/T or ADMETox (when toxicity assessment is included).

Computer-based methods have been employed in the prediction of ADMET properties of drug leads at early stages of drug discovery and such approaches are becoming increasingly popular (Lipinski et al. 1997; Lombardo et al. 2003; Gleeson et al. 2011). The rationale behind in silico approaches are the relatively lower cost and the time factor involved, when compared to standard experimental approaches for ADMET profiling (DiMasi et al. 2003; Darvas et al. 2002). As an example, it takes a minute in an in silico model to screen 20,000 molecules, but takes 20 weeks in the "wet" laboratory to do the same exercise (Hodgson 2001). Due to the accumulated ADMET data in the late 1990s, many pharmaceutical companies are now using computational models that, in some cases, are replacing the "wet" screens (Hodgson 2001). This paradigm shift has therefore spurred up the development of several theoretical methods for the prediction of ADMET parameters. A host of these theoretical models have been implemented in a number of software programs currently available for drug discovery protocols (OCHEM platform 2009; Lhasa 2010; Schrodinger 2011a; Cruciani et al. 2000), even though some of the predictions are often disappointing (Tetko et al. 2006). The software tools currently used to predict the ADMET properties of potential drug candidates often make use of quantitative structure-activity relationships, QSAR (Tetko et al. 2006; Hansch et al. 2004) or knowledge-base methods (Greene et al. 1999; Button et al. 2003; Cronin 2003). A promising lead compound may therefore be defined as one which combines potency with an admirable ADMET profile (commonly referred to as a compound's CV). As such, compounds with uninteresting predicted ADMET profiles may be completely dismissed from the list of potential drug candidates (even if these prove to be highly potent). Otherwise,

the DMPK properties are "fine-tuned" in order to improve their chances of making it to clinical trials (Hou and Wang 2008). This may explain why the "graveyard" of very highly potent compounds which do not make it to clinical trials keeps filling up, to the extent that experts in drug discovery are often faced with the challenge of either resorting to new lead compounds or "resurrecting" some buried leads with the view of "fine-tuning" their DMPK properties.

A natural product compound database built on information collected from several literature sources on medicinal plants from Central African countries, currently used in ATM, has been recently developed at our laboratory. The plants had been harvested from 10 countries (Burundi, Cameroon, Central African Republic, Chad, Congo, Equatorial Guinea, Gabon, the Democratic Republic of Congo, Rwanda and the Republic of São Tomé and Príncipe). This NP library currently contains ~3,200 compounds and preliminary analyses have proven the dataset to be sufficiently "drug-like" and diverse to be employed in lead discovery programs (Ntie-Kang et al. 2013a; Ntie-Kang et al.: ConMedNP: a natural product library from Central African medicinal plants for drug discovery. RSC Adv, submitted). Additional arguments in favour of the use of this database are the wide range of the previously observed biological activities of the compounds and the wide range of ailments being treated by traditional medicine with the help of the herbs from which the compounds have been derived (Ntie-Kang et al. 2013a, b; Zofou et al. 2013; Ntie-Kang et al.: ConMedNP: a natural product library from Central African medicinal plants for drug discovery. RSC Adv, submitted).

Since numerous drugs and many more lead compounds fail due to adverse pharmacokinetic properties at a late stage of pharmaceutical development (Darvas et al. 2002), it has become important to incorporate ADME properties' prediction into the lead compound selection early enough, by means of molecular descriptors. A molecular descriptor may be defined as a structural or physico-chemical property of a molecule or part of a molecule, for example logarithm of the n-octanol/water partition coefficient (log P), the molar weight (MW) and the total polar surface area (TPSA). A number of relevant molecular properties (descriptors) are often used to help in the assessment of the DMPK properties of potential drug leads. In this paper, an attempt has been made to carry out an in silico assessment of the ADMET profile of this dataset. A number of computed molecular descriptors, currently implemented in a wide range of software, have been used as indicators of the pharmacokinetic properties of a large proportion of currently known drugs.

Methods

Data sources and generation of 3D structures

The plant sources, geographical collection sites, chemical structures of pure compounds as well as their measured biological activities, were retrieved from literature sources and have been previously described (Ntie-Kang et al. 2013b; Zofou et al. 2013). The 3D structures of the compounds had been sketched and energy minimisation subsequently carried out using a previously described protocol (Ntie-Kang et al. 2013a).

Initial treatment of chemical structures and calculation of ADMET-related descriptors

The 3,179 low energy 3D chemical structures in the ConMedNP library were saved in .mol2 format and initially treated with LigPrep (Schrödinger 2011a). This implementation was carried out with the graphical user interface (GUI) of the Maestro software package (Schrödinger 2011b), using the Optimized Potentials for Liquid Simulations (OPLS) forcefield (Shivakumar et al. 2010; Jorgensen et al. 1996; Jorgensen et al. 1988). Protonation states at biologically relevant pH were correctly assigned (group I metals in simple salts were disconnected, strong acids were deprotonated, strong bases protonated, while topological duplicates and explicit hydrogens were added). All molecular modelling was carried out on a Linux workstation with a 3.5 GHz Intel Core2 Duo processor. A set of ADMET-related properties (a total of 46 molecular descriptors) were calculated by using the QikProp program (Schrödinger 2011c)

running in normal mode. QikProp generates physically relevant descriptors, and uses them to perform ADMET predictions. An overall ADME-compliance score – drug-likeness parameter (indicated by #stars), was used to assess the pharmacokinetic profiles of the compounds within the ConMedNP library. The #stars parameter indicates the number of property descriptors computed by QikProp that fall outside the optimum range of values for 95% of known drugs. The methods implemented were developed by Jorgensen and Duffy (Jorgensen and Duffy 2002; Duffy and Jorgensen 2000; Jorgensen and Duffy 2000). Some of the computed ADMET descriptors are shown in Table 1, along with their recommended ranges for 95% of known drugs.
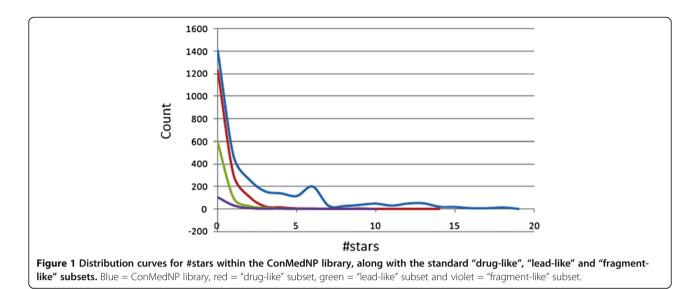
Results and discussion

Overall DMPK compliance of the ConMedNP library

The 24 most relevant molecular descriptors calculated by QikProp are used to determine the #star parameter (Schrödinger 2011d). A plot of the #stars parameter (on x-axis) against the corresponding counts (on y-axis) in the ConMedNP is plotted within the same set of axes with those of the "drug-like", "lead-like", and "fragment-like" standard subsets (Figure 1). The criteria for the respective standard subsets were defined as (MW < 500; log P < 5; HBD ≤ 5; HBA ≤ 10) (Lipinski et al. 1997), (150 ≤ MW ≤ 350; log P ≤ 4; HBD ≤ 3; HBA ≤ 6) (Teague et al. 1999; Oprea 2002; Schneider 2002) and (MW ≤ 250; -2 ≤ log P ≤ 3; HBD < 3; HBA < 6; NRB < 3) (Verdonk et al. 2003). The ADMET

Table 1 Selected computed ADMET-related descriptors and their recommended ranges for 95% of known drugs

Property	Description	Recommended range
S_{mol}	the total solvent-accessible molecular surface, in Å² (probe radius 1.4 Å)	300 to 1000 Å²
$S_{mol,hfob}$	the hydrophobic portion of the solvent-accessible molecular surface, in Å² (probe radius 1.4 Å)	0 to 750 Å²
V_{mol}	the total volume of molecule enclosed by solvent-accessible molecular surface, in Å³ (probe radius 1.4 Å)	500 to 2000 Å³
log S_{wat}	the logarithm of aqueous solubility (Jorgensen and Duffy 2002; Jorgensen and Duffy 2000)	−6.0 to 0.5
log K_{HSA}	the logarithm of predicted binding constant to human serum albumin (Colmenarejo 2001)	−1.5 to 1.2
log B/B	the logarithm of predicted blood/brain barrier partition coefficient (Luco 1999; Kelder et al. 1999; Ajay et al. 1999)	−3.0 to 1.0
BIP_{caco-2}	the predicted apparent Caco-2 cell membrane permeability, in nm s⁻¹ (in Boehringer–Ingelheim scale, Yazdanian et al. Yazdanian et al. 1998; Irvine et al. 1999; Stenberg 2001)	< 5 low, > 100 high
$MDCK$	the predicted apparent Madin-Darby canine kidney cell permeability in nm s⁻¹ (Irvine et al. 1999)	< 25 poor, > 500 great
Ind_{coh}	the index of cohesion interaction in solids, calculated from the number of hydrogen bond acceptors (HBA), donors (HBD) and the surface area accessible to the solvent, SASA (S_{mol}) by the relation $\mathbf{Ind_{coh}} = \mathbf{HBA} \times \frac{\sqrt{\mathbf{HBD}}}{\mathbf{S_{mol}}}$ (Jorgensen and Duffy 2000)	0.0 to 0.05
Glob	the globularity descriptor, Glob = $(4\pi r^2)/S_{mol}$, where r is the radius of the sphere whose volume is equal to the molecular volume	0.75 to 0.95
QP_{polrz}	the predicted polarizability	13.0 to 70.0
log $HERG$	the predicted IC$_{50}$ value for blockage of HERG K⁺ channels, (Cavalli et al. 2002; De Ponti et al. 2001)	concern < −5
log K_p	the predicted skin permeability (Potts and Guy 1992; Potts and Guy 1995)	−8.0 to −1.0
#metab	the number of likely metabolic reactions	1 to 8

Figure 1 Distribution curves for #stars within the ConMedNP library, along with the standard "drug-like", "lead-like" and "fragment-like" subsets. Blue = ConMedNP library, red = "drug-like" subset, green = "lead-like" subset and violet = "fragment-like" subset.

descriptors for some 67 compounds in the total library were not computed by QikProp, probably due to some technical details related to the working of the software which was beyond our notice. Of the remaining 3,112 compounds, 45.31% showed #star = 0, while 68.93% had #star ≤ 2. Among the 1,696 compounds of the "drug-like" subset whose pharmacokinetic properties were predicted,

72.52% had pharmacokinetic descriptors within the acceptable range for 95% of known drugs, while 96.88% showed #stars ≤ 2. The "lead-like" and "fragment-like" subsets were respectively 80.68% and 65.58% compliant for all of the 24 most relevant computed descriptors. The mean values for 19 selected computed descriptors have been shown in Table 2 for all 4 compound libraries, while

Table 2 Summary of mean pharmacokinetic property distributions of the total ConMedNP library in comparison with the various subsets

Library name	[a]Lib. size	[b]No. compl.	[c]MW (Da)	[d]LogP	[e]HBA	[f]HBD	[g]NRB
Total library	3,118	1,410	421.72	3.86	6.54	2.03	7.46
Drug-like	1,696	1,230	326.14	2.76	5.13	1.43	4.47
Lead-like	730	589	269.55	2.24	4.26	1.17	3.23
Fragment-like	154	101	192.98	1.33	3.58	0.89	2.17
	[h]LogB/B	[i]BIP$_{caco-2}$ (nm s^{-1})	[j]S_{mol} (Å^2)	[k]$S_{mol,hfob}$ (Å^2)	[l]V_{mol} (Å^3)	[m]LogS_{wat} (S in mol L^{-1})	[n]LogK_{HSA}
Total library	−1.23	1322.58	696.98	420.94	1314.78	−5.16	0.52
Drug-like	−0.77	1195.51	564.70	274.17	1014.47	−3.78	0.13
Lead-like	−0.57	1233.11	494.87	204.27	856.93	−3.04	−0.07
Fragment-like	−0.47	1178.35	393.88	111.25	640.36	−1.78	−0.48
	[o]MDCK	[p]Ind$_{coh}$	[q]Glob	[r]QP$_{polrz}$ (Å^3)	[s]LogHERG	[t]LogK_p	[u]# metab
Total library	765.31	0.013	0.84	42.99	−4.60	−2.99	5.53
Drug-like	689.00	0.009	0.87	33.17	−4.39	−2.91	4.55
Lead-like	717.25	0.008	0.88	27.75	−4.17	−2.85	3.40
Fragment-like	682.72	0.007	0.91	19.67	−3.46	−2.81	1.90

[a]Size or number of compounds in library; [b]Number of compounds with #star = 0; [c]Molar weight (range for 95% of drugs: 130–725 Da); [d]Logarithm of partitioning coefficient between *n*-octanol and water phases (range for 95% of drugs: -2 to 6.5); [e]Number of hydrogen bonds accepted by the molecule (range for 95% of drugs: 2–20); [f]Number of hydrogen bonds donated by the molecule (range for 95% of drugs: 0–6).; [g]Number of rotatable bonds (range for 95% of drugs: 0–15); [h]Logarithm of predicted blood/brain barrier partition coefficient (range for 95% of drugs: -3.0 to 1.0); [i]Predicted apparent Caco-2 cell membrane permeability in Boehringer–Ingelheim scale, in nm/s (range for 95% of drugs: < 5 low, > 500 high); [j]Total solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 300–1000 Å^2); [k]Hydrophobic portion of the solvent-accessible molecular surface, in Å^2 (probe radius 1.4 Å) (range for 95% of drugs: 0–750 (Å^2); [l]Total volume of molecule enclosed by solvent-accessible molecular surface, in Å^3 (probe radius 1.4 Å) (range for 95% of drugs: 500–2000 Å^3); [m]Logarithm of aqueous solubility (range for 95% of drugs: -6.0 to 0.5); [n]Logarithm of predicted binding constant to human serum albumin (range for 95% of drugs: -1.5 to 1.5); [o]Predicted apparent MDCK cell permeability in nm/sec (< 25 poor, > 500 great); [p]Index of cohesion interaction in solids (0.0 to 0.05 for 95% of drugs); [q]Globularity descriptor (0.75 to 0.95 for 95% of drugs); [r]Predicted polarizability (13.0 to 70.0 for 95% of drugs); [s]Predicted IC$_{50}$ value for blockage of HERG K$^+$ channels (concern < −5); [t]Predicted skin permeability (−8.0 to −1.0 for 95% of drugs); [u]Number of likely metabolic reactions (range for 95% of drugs: 1–8).

percentage compliances for 14 selected parameters are shown in Table 3. The mean values were used to assess the probability of finding drug leads within the ConMedNP compound library.

Bioavailability prediction

The bioavailability of a compound depends on the processes of absorption and liver first-pass metabolism (Van de Waterbeemd and Gifford 2003). Absorption in turn depends on the solubility and permeability of the compound, as well as interactions with transporters and metabolizing enzymes in the gut wall. The computed parameters used to assess oral absorption are the predicted aqueous solubility, $logS_{wat}$, the conformation-independent predicted aqueous solubility, CI $logS_{wat}$, the predicted qualitative human oral absorption, the predicted % human oral absorption and compliance to Jorgensen's famous "Rule of Three" (ro3). The solubility calculation procedure implemented depends on the similarity property space between the given molecule and its most similar analogue within the experimental training set used to develop the model implemented in QikProp, i.e., if the similarity is < 0.9, then the QikProp predicted value is taken, otherwise, the predicted property, P_{pred}, is adjusted such that:

$$P_{pred} = SP_{exp} + (1-S)P_{QP} \qquad (1)$$

where S is the similarity, and P_{exp} and P_{QP} are the respective experimental and QikProp predictions for the most similar molecule within the training set. In equation (1), if $S = 1$, then the predicted property is equal to the measured experimental property of the training set compound. According to Jorgensen's ro3, if a compound complies to all or some of the rules ($logS_{wat} > -5.7$, $BIP_{caco-2} > 22$ nm/s and # Primary Metabolites < 7), then it is more likely to be orally available. The distribution curves for two of the three
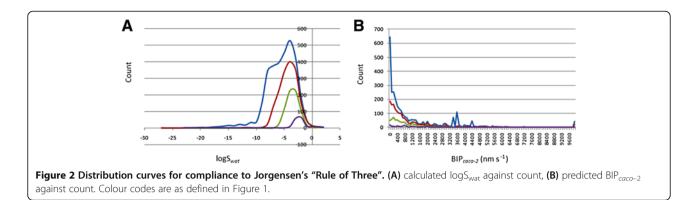
determinants for the ro3 ($logS_{wat}$ and BIP_{caco-2}) are shown in Figure 2. In general 43.57% of the ConMedNP library was compliant to the ro3, while the respective % compliances for the various subsets were 73.52%, 93.56% and 100% for the "drug-like", "lead-like" and "fragment-like" libraries. Among the individual computed parameters, the most remarkable was $logS_{wat}$. This was met by 72.46% of the compounds within the ConMedNP library, while this property showed a Gaussian distribution for the "drug-like" and "lead-like" subsets. Only 37.28% of the compounds fell within the respected range for the BIP_{caco-2} criterion. The predicted apparent Caco-2 cell permeability, BIP_{caco-2} (in nm s^{-1}), model the permeability of the gut-blood barrier (for non-active transport), even though this parameter is not often correctly predicted computationally (Veber et al. 2002). The histograms of the predicted qualitative human oral absorption parameter (in the scale 1 = low, 2 = medium and 3 = high) are shown in Figure 3. It was observed 48.65% of the compounds in ConMedNP were predicted to have high human oral absorption. The predicted % human oral absorption (on 0 to 100% scale) shows a similar trend, 42.09% of the compounds being predicted to be absorbed at 100% and 57.81% of the compounds predicted to be absorbed at > 90%.

A molecule's size, as well as its capacity to make hydrogen bonds, its overall lipophilicity and its shape and flexibility are important properties to consider when determining permeability. Molecular flexibility has been seen as a parameter which is dependent on the number of rotatable bonds (NRB), a property which influences bioavailability in rats (Veber et al. 2002). The distribution of the NRB for this dataset has been discussed in detail elsewhere (Ntie-Kang et al. ConMedNP: a natural product library from Central African medicinal plants for drug discovery. RSC Adv, submitted) and the results reveal that the compounds within the ConMedNP library show some degree of conformational

Table 3 Summary of percentage compliances of selected ADMET-related descriptors of the total ConMedNP library in comparison with the various subsets

Library name	*LogB/B	*BIP$_{caco-2}$ (nm s^{-1})	*S$_{mol}$ (Å^2)	*S$_{mol,hfob}$ (Å^2)	*V$_{mol}$ (Å^3)	*LogS$_{wat}$ (S in mol L^{-1})	*LogK$_{HSA}$
Total library	88.53	37.28	90.36	91.32	91.03	72.46	81.01
Drug-like	99.35	43.99	99.35	99.82	98.99	90.39	99.35
Lead-like	99.72	53.70	99.72	100.00	99.72	99.31	99.59
Fragment-like	100.00	33.11	95.45	100.00	92.21	97.40	98.05
	*MDCK	*Ind$_{coh}$	*Glob	^{a}ro3	*LogHERG	*LogK$_P$	*# metab
Total library	47.14	94.92	89.88	43.57	58.35	92.09	81.52
Drug-like	59.61	99.14	97.70	73.52	63.62	95.93	91.89
Lead-like	59.86	100.00	97.81	93.56	76.03	97.53	96.44
Fragment-like	63.33	100.00	92.21	100.00	100.00	98.05	92.21

*Descriptors are defined in Tables 1 and 2; ^{a}percentage compliance to Jorgensen's Rule of Three.

Figure 2 Distribution curves for compliance to Jorgensen's "Rule of Three". (A) calculated $logS_{wat}$ against count, **(B)** predicted BIP_{caco-2} against count. Colour codes are as defined in Figure 1.

flexibility, the peak value for the NRB being between 1 and 2, while the average value is 7.46 (Table 2).
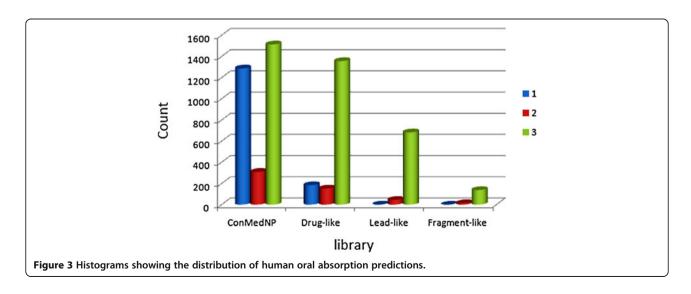
Prediction of blood–brain barrier (BBB) penetration

Too polar drugs do not cross the BBB. The blood/brain partition coefficients (logB/B) were computed and used as a predictor for access to the central nervous system (CNS). The predicted CNS activity was computed on a −2 (inactive) to +2 (active) scale and showed that only 2.47% of the compounds in ConMedNP could be active in the CNS (predicted CNS activity > 1). A distribution of logB/B (Figure 4) shows a right-slanted Gaussian-shaped curve with a peak at −0.5 logB/B units (the same for all the standard subsets), with 88.53% of the compounds in ConMedNP falling within the recommended range for the predicted brain/blood partition coefficient (−3.0 to 1.2). Madin-Darby canine kidney (MDCK) monolayers, are widely used to make oral absorption estimates, the reason being that these cells also express transporter proteins, but only express very low levels of metabolizing enzymes (Veber et al. 2002). They are also used as an additional criterion to predict BBB penetration. Thus, our calculated apparent MDCK cell permeability could be considered to be a good mimic for the BBB (for non-active

transport). It was estimated that only about 47% of the compounds had apparent MDCK cell permeabilities which fall within the recommended range of 25–500 nm s⁻¹ for 95% of known drugs. This situation knew improvements in the "drug-like" and "lead-like" subsets (~60% for both subsets).
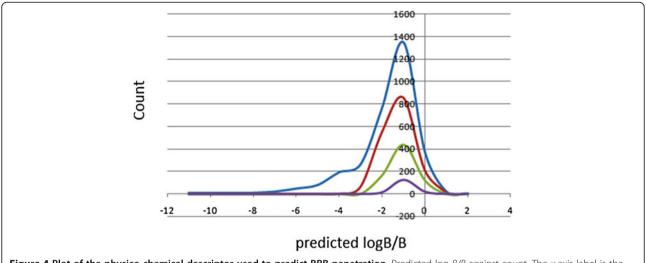
Prediction of dermal penetration

This factor is important for drugs administered through the skin. The distribution of computed skin permeability parameter, $\log K_p$, showed smooth Gaussian-shaped graphs centred at −2.5 $\log K_P$ units for all 4 datasets (Figure 5), with ~92% of the compounds in the dataset falling within the recommended range for 95% of known drugs. The predicted maximum transdermal transport rates, J_m (in μ cm⁻² hr⁻¹), were computed from the aqueous solubility (S_{wat}) and skin permeability (K_p), using the relation (2):

$$J_m = K_P \times MW \times S_{wat} \tag{2}$$

This parameter showed variations from 0 to 1603 μ cm⁻² hr⁻¹, with only about 1.38% of the compounds in ConMedNP having predicted value of $J_m > 100$ μ cm⁻² hr⁻¹.

Figure 3 Histograms showing the distribution of human oral absorption predictions.

Figure 4 Plot of the physico-chemical descriptor used to predict BBB penetration. Predicted log *B/B* against count. The *x*-axis label is the lower limit of binned data, e.g. 0 is equivalent to 0.0 to 1.0. Colour codes are as defined in Figure 1.
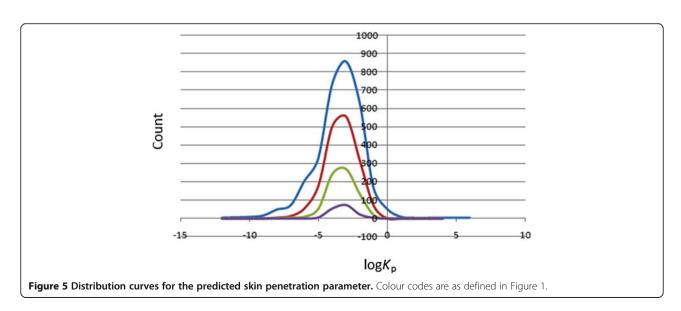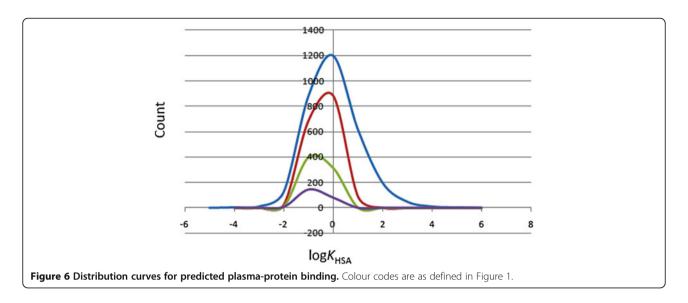
Prediction of plasma-protein binding

The efficiency of a drug may be affected by the degree to which it binds to the proteins within blood plasma. It is noteworthy that binding of drugs to plasma proteins (like human serum albumin, lipoprotein, glycoprotein, α, β, and γ globulins) greatly reduces the quantity of the drug in general blood circulation and hence the less bound a drug is, the more efficiently it can traverse cell membranes or diffuse. The predicted plasma-protein binding has been estimated by the prediction of binding to human serum albumin; the log K_{HSA} parameter (recommended range is −1.5 to 1.5 for 95% of known drugs). Figure 6 shows the variation of this calculated parameter within the ConMedNP dataset, as well as for the standard subsets. This equally gave smooth Gaussian-shaped curves centred on −0.5 log K_{HSA} units

for the total and "drug-like" libraries and −1.5 log K_{HSA} units for the "lead-like" and "fragment-like" datasets. In addition, our calculations reveal that > 81% of the compounds within the ConMedNP library are compliant to this parameter, indicating that a majority of the compounds are likely to circulate freely within the blood stream and hence have access to the target site.

Metabolism prediction

An estimated number of possible metabolic reactions has also been predicted by QikProp and used to determine whether the molecules can easily gain access to the target site after entering the blood stream. The average estimated number of possible metabolic reactions for the ConMedNP library is between 5 and 6,

Figure 5 Distribution curves for the predicted skin penetration parameter. Colour codes are as defined in Figure 1.

Figure 6 Distribution curves for predicted plasma-protein binding. Colour codes are as defined in Figure 1.
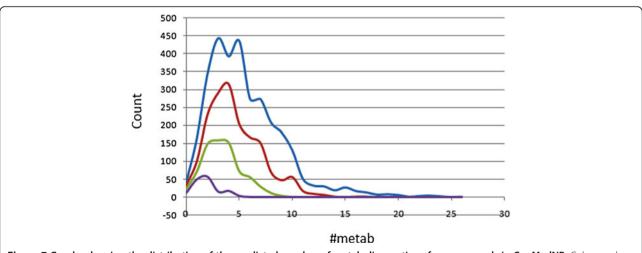
while those of the standard subsets are respectively between 4 and 5, between 3 and 4 and between 1 and 2 for the "drug-like", "lead-like" and "fragment-like" libraries (Table 2). Even though some of the compounds are likely to undergo as many as up to 26 metabolic reactions due to the complexity of some of the plant secondary metabolites within the database (Figure 7), ~81% of the compounds are predicted to undergo the recommended number of metabolic steps (1 to 8 reactions), with the situation improving to ~92% and almost 100% in the "drug-like" and "lead-like" subsets respectively. From Figure 7, it can be observed that the total and "lead-like" libraries show peak values at 3 metabolic steps, while the drug-like subset rather shows a peak at 4 metabolic steps and the "fragment-like" subsets peaks at 2 predicted metabolic reactions.

Prediction of blockage of human ether-a-go-go-related gene potassium (HERG K+) channel

Human ether-a-go-go related gene (HERG) encodes a potassium ion (K+) channel that is implicated in the fatal arrhythmia known as *torsade de pointes* or the long QT syndrome (Hedley et al. 2009). The HERG K+ channel, which is best known for its contribution to the electrical activity of the heart that coordinates the heart's beating, appears to be the molecular target responsible for the cardiac toxicity of a wide range of therapeutic drugs (Vandenberg 2001). HERG has also been associated with modulating the functions of some cells of the nervous system and with establishing and maintaining cancer-like features in leukemic cells (Chiesa et al. 1997). Thus, HERG K+ channel blockers are potentially toxic and the predicted IC_{50} values often provide reasonable predictions for cardiac toxicity of drugs in the early stages of

Figure 7 Graphs showing the distribution of the predicted number of metabolic reactions for compounds in ConMedNP. Colour codes are as defined in Figure 1.

drug discovery (Aronov 2005). In this work, the estimated or predicted IC_{50} values for blockage of this channel have been used to model the process *in silico*. The recommended range for predicted log IC_{50} values for blockage of HERG K^+ channels (logHERG) is > –5. A distribution curve for the variation of the predicted logHERG is shown in Figure 8, which is a left-slanted Gaussian-shaped curve, with a peak at –5.5 logHERG units for the total library, as well as for the "drug-like" and "lead-like" subsets. It was observed that in general, this parameter is predicted to fall within the recommended range for about 58% of the compounds within the ConMedNP dataset, ~64% for the "drug-like" subset and ~76% for the "lead-like" subset.
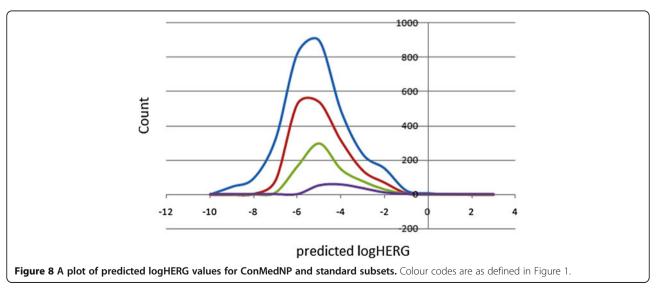
Usefulness of the compound library

The usefulness of the ConMedNP database in lead generation has been exemplified with the docking and pharmacophore-based screening for potential inhibitors of a validated anti-malarial drug target in our laboratory, and the results will be published in a subsequent paper. It is important to mention that virtual screening results could provide insight and direct natural product chemists to search for theoretically active principles with attractive ADMET profiles, which have been previously isolated, but not tested for activity against specified drug targets (if samples are absent). This "resurrection" process could prove to be a better procedure for lead search than the random screening, which is a common practice in our African laboratories. This dataset is constantly being updated; meanwhile a MySQL platform to facilitate the searching of this database and ordering of compound samples is under development within our group and will also be published subsequently. However, 3D structures of the compounds, as well as their physico-chemical properties that were used to evaluate

the DMPK profile, can be freely downloaded as additional files accompanying this publication (Additional files 1, 2, 3 and 4). In addition, information about compound sample availability can be obtained on request from the authors of this paper or from the pan-African Natural Products Library (p-ANAPL) project (Chibale et al. 2012; p-ANAPL 2013).

Conclusions

Modern drug discovery programs usually involve the search for small molecule leads with attractive pharmacokinetic profiles. The presence of such within the ConMedNP library is of major importance and therefore renders the database attractive, in addition to the already known properties – "drug-like", "lead-like", "fragment-like" and diverse. This is an indication that the 3D structures of naturally occurring compounds within ConMedNP could be a good starting point for docking, neural networking and pharmacophore-based virtual screening campaigns, thus rendering ConMedNP a useful asset for the drug discovery community.

Availability and requirements

3D structures of the compounds, as well as their physico-chemical properties that were used to evaluate the DMPK profile of the ConMedNP library, can be freely downloaded (for non commercial use) as additional files which accompany this publication (Additional files 1, 2, 3 and 4). Physical samples for testing are available at the various research laboratories in Central Africa in varying quantities. Questions regarding the available of compound samples could be addressed directly to the authors of this paper. Otherwise samples could be obtainable from the p-ANAPL consortium, which has a mandate to collect samples of NPs from the entire continent of Africa

Figure 8 A plot of predicted logHERG values for ConMedNP and standard subsets. Colour codes are as defined in Figure 1.

and make them available for biological screening. This network is being set up under the auspices of the Network for Analytical and Bioassay Services in Africa (NABSA) (Chibale et al. 2012; p- ANAPL 2013).

Additional files

> **Additional file 1: 3D structures of compounds currently included in ConMedNP with calculated pharmacokinetic descriptors.**
>
> **Additional file 2: 3D structures of the "drug-like" subset with calculated pharmacokinetic descriptors.**
>
> **Additional file 3: 3D structures of the "lead-like" subset with calculated pharmacokinetic descriptors.**
>
> **Additional file 4: 3D structures of the "fragment-like" subset with calculated pharmacokinetic descriptors.**

Abbreviations

3D: Three dimensional; ADME/T: Absorption, distribution, metabolism, excretion, and toxicity; CADD: Computer-aided drug design; ConMedNP: Congo basin medicinal plant and natural products database; DMPK: Drug metabolism and pharmacokinetics; log P: Logarithm of the octan-1-ol/water partition coefficient; MDCK: Madin-Darby canine kidney; MW: Molar weight; NABSA: Network for analytical and bioassay services in Africa; NP: Natural product; NRB: Number of rotatable bonds; OPLS: Optimized potentials for liquid simulations; p-ANAPL: pan-African Natural Products Library; SASA: Solvent accessible surface area; TPSA: Total polar surface area; VS: Virtual screening.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WS, EM, LCOO and SMNE conceived the project. All authors participated in the data generation and analysis, discussion of results and the conception of the paper. FNK wrote the first draft of the paper and all authors agreed on the final version before submission. This work is part of the PhD project of FNK.

Authors' informations

WS and SMNE are professors of medicinal chemistry with an interest in CADD, while SMNE also focuses organic synthesis and on natural product leads from Cameroonian medicinal plants. LMM and JAM are natural product chemists actively involved in the isolation and characterization of secondary metabolites from Cameroonian medicinal plants. LLL holds a PhD in environmental science and manages a Chemical and Bioactivity Information Centre (CBIC) with a focus on developing databases for information from medicinal herbs in Africa. PNJ is a retired research officer of Lhasa Ltd who currently leads the CBIC branch in Leeds, UK. FNK is a PhD student working on CADD under the joint supervision of LCOO and EM.

Author details

[1]CEPAMOQ, Faculty of Science, University of Douala, P.O. Box 8580, Douala, Cameroon. [2]Chemical and Bioactivity Information Centre, Department of Chemistry, Faculty of Science, University of Buea, P.O. Box 63, Buea, Cameroon. [3]Department of Pharmaceutical Sciences, Martin-Luther University of Halle-Wittenberg, Wolfgang-Langenbeck Str. 4, 06120, Halle (Saale), Germany. [4]Department of Chemistry, Faculty of Science, University of Buea, P.O. Box 63, Buea, Cameroon. [5]Laboratory for Simulations and Biomolecular Physics, Advanced Teachers Training College, University of Yaoundé I, P.O. Box 47, Yaoundé, Cameroon. [6]Laboratory of Fundamental and Applied Physics, University of Abobo-Adjame, Abidjan 02 BP 801, Cote d'Ivoire. [7]Department of Chemistry, Faculty of Science, University of Douala, P. O. Box 24157, Douala, Cameroon. [8]Chemical and Bioactivity Information Centre, 22-23 Blenheim Terrace, Woodhouse Lane, Leeds LS2 9HD, UK.

References

Ajay, Bermis GW, Murkco MA (1999) Designing libraries with CNS activity. J Med Chem 42:4942–4951

Aronov AM (2005) Predictive in silico modeling for hERG channel blockers. Drug Discov Today 10:149–155

Butler MS (2005) Natural products to drugs: natural product derived compounds in clinical trials. Nat Prod Rep 22:162–195

Button WG, Judson PN, Long A, Vessey JD (2003) Using absolute and relative reasoning in the prediction of the potential metabolism of xenobiotics. J Chem Inf Comput Sci 43:1371–1377

Cavalli A, Poluzzi E, De Ponti F, Recanatini M (2002) Toward a pharmacophore for drugs inducing the long QT syndrome: insights from a CoMFA study of HERG K+ channel blockers. J Med Chem 45:3844–3853

Chibale K, Davies-Coleman M, Masimirembwa C (2012) Drug discovery in Africa: impacts of genomics, natural products, traditional medicines, insights into medicinal chemistry, and technology platforms in pursuit of new drugs. Springer, Berlin

Chiesa N, Rosati B, Arcangeli A, Olivotto M, Wanke E (1997) A novel role for HERG K+ channels: spike-frequency adaptation. J Physiol 501:313–318

Chin YW, Balunas MJ, Chai HB, Kinghorn AD (2006) Drug discovery from natural sources. AAPS J 8(2):E239–E253

Colmenarejo G, Alvarez-Pedraglio A, Lavandera J-L (2001) Cheminformatic models to predict binding affinities to human serum albumin. J Med Chem 44:4370–4378

Cronin MTD (2003) Computer-assisted prediction of drug toxicity and metabolism in modern methods of drug discovery. In: Hilgenfeld R, Hillisch A (eds) Modern methods of drug discovery. Birkhäuser, Basel

Cruciani C, Crivori P, Carrupt PA, Testa B (2000) Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. J Mol Struc-Theochem 503:17–30

Darvas F, Keseru G, Papp A, Dormán G, Urge L, Krajcsi P (2002) In Silico and Ex silico ADME approaches for drug discovery. Top Med Chem 2:1287–1304

De Ponti F, Poluzzi E, Montanaro N (2001) Organising evidence on QT prolongation and occurrence of Torsades de Pointes with non-antiarrhythmic drugs: a call for consensus. Eur J Clin Pharmacol 57:185–209

DiMasi JA, Hansen RW, Grabowsk HG (2003) The price of innovation: new estimates of drug development costs. J Health Econ 22:151–185

Duffy EM, Jorgensen WL (2000) Prediction of properties from simulations: free energies of solvation in hexadecane, octanol, and water. J Am Chem Soc 122:2878–2888

Efange SMN (2002) Natural products: a continuing source of inspiration for the medicinal chemist. In: Iwu MM, Wootton JC (eds) Advances in phytomedicine. Elsevier Science, Amsterdam

Gleeson MP, Hersey A, Hannongbua S (2011) In-silico ADME models: a general assessment of their utility in drug discovery applications. Curr Top Med Chem 11(4):358–381

Grabowski K, Schneider G (2007) Properties and architecture of drugs and natural products revisited. Curr Chem Biol 1:115–127

Grabowski K, Baringhaus K-H, Schneider G (2008) Scaffold diversity of natural products: inspiration for combinatorial library design. Nat Prod Rep 25:892–904

Greene J, Judson PN, Langowski JJ (1999) Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. SAR QSAR Environ Res 10:299–314

Hansch C, Leo A, Mekapatia SB, Kurup A (2004) QSAR and ADME. Bioorg Med Chem 12:3391–3400

Harvey AL (2008) Natural products in drug discovery. Drug Discov Today 13:894–901

Hedley PL, Jørgensen P, Schlamowitz S, Wangari R, Moolman-Smook J, Brink PA, Kanters JK, Corfield VA, Christiansen M (2009) The genetic basis of long QT and short QT syndromes: a mutation update. Human Mutation 30:1486–1511

Hodgson J (2001) ADMET – turning chemicals into drugs. Nat Biotechnol 19:722–726

Hou T, Wang J (2008) Structure-ADME relationship: still a long way to go? Expert Opin Drug Metab Toxicol 4(6):759–770

Irvine JD, Takahashi L, Lockhart K, Cheong J, Tolan JW, Selick HE, Grove JR (1999) MDCK (Madin-Darby canine kidney) cells: a tool for membrane permeability screening. J Pharm Sci 88:28–33

Jorgensen WL, Duffy EM (2000) Prediction of drug solubility from Monte Carlo simulations. Bioorg Med Chem Lett 10:1155–1158

Jorgensen WL, Duffy EM (2002) Prediction of drug solubility from structure. Adv Drug Deliv Rev 54:355–366

Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 110(6):1657–1666

Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236

Kelder J, Grootenhuis PD, Bayada DM, Delbresine LP, Ploemen JP (1999) Polar molecular surface as a dominating determinant for oral absorption and brain pernetration of drugs. Pharm Res 16:1514–1519

Koehn FE, Carter GT (2005) The evolving role of natural products in drug discovery. Nat Rev Drug Discov 4:206–220

Lhasa Ltd (2010) Meteor, version 13.0.0. Lhasa, Leeds, UK

Li JWH, Vederas JC (2009) Drug discovery and natural products: end of an era or an endless frontier? Science 325:161–165

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 23:3–25

Lombardo F, Gifford E, Shalaeva MY (2003) *In silico* ADME prediction: data, models, facts and myths. Mini Rev Med Chem 3:861–875

Luco JM (1999) Prediction of brain–blood distribution of a large set of drugs from structurally derived descriptors using partial least squares (PLS) modelling. J Chem Inf Comput Sci 39:396–404

Navia MA, Chaturvedi PR (1996) Design principles for orally bioavailable drugs. Drug Dev Today 1:179–189

Newman DJ (2008) Natural products as leads to potential drugs: an old process or the new hope for drug discovery? J Med Chem 51:2589–2599

Ntie-Kang F, Mbah JA, Mbaze LM, Lifongo LL, Scharfe M, Ngo Hanna J, Cho-Ngwa F, Amoa Onguéné P, Owono Owono LC, Megnassan E, Sippl W, Efange SMN (2013a) CamMedNP: Building the Cameroonian 3D structural natural products database for virtual screening. BMC Complement Altern Med 13:88

Ntie-Kang F, Lifongo LL, Mbaze LM, Ekwelle N, Owono Owono LC, Megnassan E, Judson PN, Sippl W, Efange SMN (2013b) Cameroonian medicinal plants: a bioactivity versus ethnobotanical survey and chemotaxonomic classification. BMC Complement Altern Med 13:147

OCHEM (2009) A platform for the creation of *in silico* ADME/Tox prediction models., http://www.eadmet.com/en/ochem.php

Oprea TI (2002) Current trends in lead discovery: are we looking for the appropriate properties? J Comput-Aided Mol Des 16:325–334

p-ANAPL (2013) pan-ANAPL: pan-African natural products library. http://www.linkedin.com/groups/pANPL-4098579/about

Potts RO, Guy RH (1992) Skin permeability. Pharm Res 9:663–669

Potts RO, Guy RH (1995) A predictive algorithm for skin permeability: the effects of molecular size and hydrogen bond activity. Pharm Res 12:1628–1633

Quinn RJ, Carroll AR, Pham MB, Baron P, Palframan ME, Suraweera L, Pierens GK, Muresan S (2008) Developing a drug-like natural product library. J Nat Prod 71:464–468

Schneider G (2002) Trends in virtual computational library design. Curr Med Chem 9:2095–2102

Schrödinger (2011a) QikProp, version 3.4. LLC, New York, NY

Schrödinger (2011b) LigPrep software, version 2.5. LLC, New York, NY

Schrödinger (2011c) Maestro, version 9.2. LLC, New York, NY

Schrödinger Press (2011) QikProp 3.4 User Manual. LLC, New York, NY

Shivakumar D, Williams J, Wu Y, Damm W, Shelley J, Sherman W (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J Chem Theory Comput 6:1509–1519

Stenberg P, Norinder U, Luthman K, Artursson P (2001) Experimental and computational screening models for the prediction of intestinal drug absorption. J Med Chem 44:1927–1937

Teague SJ, Davis AM, Leeson PD, Opea TI (1999) The design of leadlike combinatorial libraries. Angew Chem, Int Ed 38:3743–3748

Tetko IV, Bruneau P, Mewes H-W, Rohrer DC, Poda GI (2006) Can we estimate the accuracy of ADMET predictions? Drug Discov Today 11:700–707

Van de Waterbeemd H, Gifford E (2003) ADMET *in silico* modelling: towards prediction paradise? Nat Rev Drug Discov 2:192–204

Vandenberg JI, Walker BD, Campbell TJ (2001) HERG K$^+$ channels: friend or foe. Trends Pharmacol Sci 22:240–246

Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem 45:2615–2623

Verdonk ML, Cole JC, Hartshorn ML, Murray CW, Taylor RD (2003) Improved protein-ligand docking using GOLD. Proteins 52:609–623

Wetzel S, Schuffenhauer A, Roggo S, Ertl P, Waldmann H (2007) Cheminformatic analysis of natural products and their chemical space. Chimia Int J Chem 61:355–360

Yazdanian M, Glynn SL, Wright JL, Hawi A (1998) Correlating partitioning and caco-2 cell permeability of structurally diverse small molecular weight compounds. Pharm Res 15:1490–1494

Zofou D, Ntie-Kang F, Sippl W, Efange SMN (2013) Bioactive natural products derived from the Central African flora against neglected tropical diseases and HIV. Nat Prod Rep 30:1098–1120